Analysis of flight landings

Pruthvi Ranjan Reddy Pati

1.Data Preparation

2. Predicting whether the landing of a flight as safe, long and risky

3. Identifying important factors that affect the landing

4. Predicting the number of passengers using Poisson regression

5. Identifying significant variables to predict number of passengers

## 1.Data Preparation

FAA\_1 <- read\_excel("FAA1(1).xls")  
FAA\_2 <- read\_excel("FAA2(1).xls")  
FAA\_2 <- mutate(FAA\_2, duration = NA)  
FAA <- rbind(FAA\_1, FAA\_2)  
FAA <- FAA[,c(1,3,4,6,7,8,2,5)]  
FAA <- FAA %>%   
 filter(duration > 40 | is.na(duration)) %>%   
 filter(between(speed\_ground, 30, 140)) %>%   
 filter(between(speed\_air, 30, 140) | is.na(speed\_air)) %>%   
 filter(height > 6)  
FAA <- arrange\_(FAA, .dots = colnames(FAA)[1:6])  
FAA <- FAA[!duplicated(FAA[,1:6]),]  
FAA <- FAA %>%   
 mutate(aircraft\_01 = case\_when(  
 FAA$aircraft == "airbus" ~ 0,  
 TRUE ~ 1))  
FAA$aircraft\_01 <- as.factor(FAA$aircraft\_01)  
faa\_multi<-FAA  
faa\_multi$y<-ifelse(faa\_multi$distance<1000,1,  
 ifelse(faa\_multi$distance>=2500,3,2))  
faa\_multi<-faa\_multi[,-c(1,6)]  
faa\_multi$y<-as.factor(faa\_multi$y)  
summary(faa\_multi)

## no\_pasg speed\_ground height pitch   
## Min. :29.00 Min. : 33.57 Min. : 6.228 Min. :2.284   
## 1st Qu.:55.00 1st Qu.: 66.19 1st Qu.:23.530 1st Qu.:3.641   
## Median :60.00 Median : 79.79 Median :30.167 Median :4.001   
## Mean :60.06 Mean : 79.59 Mean :30.472 Mean :4.005   
## 3rd Qu.:65.00 3rd Qu.: 91.95 3rd Qu.:37.014 3rd Qu.:4.369   
## Max. :87.00 Max. :136.66 Max. :59.946 Max. :5.927   
##   
## duration speed\_air aircraft\_01 y   
## Min. : 41.95 Min. : 90.00 0:444 1:269   
## 1st Qu.:119.65 1st Qu.: 96.25 1:389 2:460   
## Median :154.26 Median :101.15 3:104   
## Mean :154.73 Mean :103.65   
## 3rd Qu.:189.64 3rd Qu.:109.40   
## Max. :305.62 Max. :136.42   
## NA's :51 NA's :629

## 

## 2.Classfying a landing into safe, long and risky

## Multinomial Regression

multi\_reg<-multinom(formula= y ~ duration+aircraft\_01+no\_pasg+ speed\_ground+height+pitch, data=faa\_multi)

## # weights: 24 (14 variable)  
## initial value 859.114810   
## iter 10 value 528.819695  
## iter 20 value 215.716132  
## iter 30 value 199.808289  
## iter 40 value 199.411638  
## iter 50 value 199.009558  
## final value 198.748962   
## converged

summary(multi\_reg)

## Call:  
## multinom(formula = y ~ duration + aircraft\_01 + no\_pasg + speed\_ground +   
## height + pitch, data = faa\_multi)  
##   
## Coefficients:  
## (Intercept) duration aircraft\_011 no\_pasg speed\_ground height  
## 2 -20.09861 -0.003528351 4.084567 -0.01864553 0.2444096 0.1564506  
## 3 -134.94460 0.001835523 9.066750 -0.01118994 1.2236524 0.3909240  
## pitch  
## 2 -0.4055071  
## 3 0.8773116  
##   
## Std. Errors:  
## (Intercept) duration aircraft\_011 no\_pasg speed\_ground height  
## 2 2.3316519 0.002795550 0.4370683 0.01790936 0.02045743 0.01860190  
## 3 0.0401298 0.008101223 0.8816893 0.05827181 0.04073377 0.04903569  
## pitch  
## 2 0.2798750  
## 3 0.7670542  
##   
## Residual Deviance: 397.4979   
## AIC: 425.4979

* Performing step regression after mean imputing duration

faa\_multi$duration<-ifelse(is.na(faa\_multi$duration),mean(faa\_multi$duration,na.rm=TRUE),faa\_multi$duration)  
  
multi\_reg1<-multinom(formula= y ~ duration+aircraft\_01+no\_pasg+ speed\_ground+height+pitch, data=faa\_multi)

## # weights: 24 (14 variable)  
## initial value 915.144036   
## iter 10 value 540.176790  
## iter 20 value 231.202800  
## iter 30 value 212.595168  
## iter 40 value 212.199624  
## iter 50 value 211.836438  
## final value 211.484876   
## converged

step\_multi<-step(multi\_reg1)

## Start: AIC=450.97  
## y ~ duration + aircraft\_01 + no\_pasg + speed\_ground + height +   
## pitch  
##   
## trying - duration   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 548.656090  
## iter 20 value 230.536949  
## iter 30 value 215.475052  
## iter 40 value 214.009972  
## iter 50 value 212.521383  
## iter 50 value 212.521383  
## iter 50 value 212.521383  
## final value 212.521383   
## converged  
## trying - aircraft\_01   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 559.456522  
## iter 20 value 315.060565  
## iter 30 value 310.576362  
## iter 40 value 310.558696  
## final value 310.557158   
## converged  
## trying - no\_pasg   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 554.198020  
## iter 20 value 231.041792  
## iter 30 value 217.924374  
## iter 40 value 215.681441  
## iter 50 value 212.510256  
## final value 212.510249   
## converged  
## trying - speed\_ground   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 766.795123  
## final value 754.882742   
## converged  
## trying - height   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 515.711181  
## iter 20 value 288.494114  
## iter 30 value 278.051516  
## iter 40 value 277.979828  
## final value 277.974787   
## converged  
## trying - pitch   
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 543.145607  
## iter 20 value 230.301156  
## iter 30 value 216.484552  
## iter 40 value 215.160323  
## final value 213.732971   
## converged  
## Df AIC  
## - no\_pasg 12 449.0205  
## - duration 12 449.0428  
## <none> 14 450.9698  
## - pitch 12 451.4659  
## - height 12 579.9496  
## - aircraft\_01 12 645.1143  
## - speed\_ground 12 1533.7655  
## # weights: 21 (12 variable)  
## initial value 915.144036   
## iter 10 value 554.198020  
## iter 20 value 231.041792  
## iter 30 value 217.924374  
## iter 40 value 215.681441  
## iter 50 value 212.510256  
## final value 212.510249   
## converged  
##   
## Step: AIC=449.02  
## y ~ duration + aircraft\_01 + speed\_ground + height + pitch  
##   
## trying - duration   
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 392.072308  
## iter 20 value 227.605114  
## iter 30 value 218.318199  
## iter 40 value 213.555421  
## final value 213.544013   
## converged  
## trying - aircraft\_01   
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 478.836303  
## iter 20 value 312.952689  
## iter 30 value 311.582863  
## iter 40 value 311.420205  
## final value 311.420199   
## converged  
## trying - speed\_ground   
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 760.209795  
## final value 755.205890   
## converged  
## trying - height   
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 419.516555  
## iter 20 value 281.322943  
## iter 30 value 279.002436  
## iter 40 value 278.763213  
## iter 40 value 278.763213  
## iter 40 value 278.763213  
## final value 278.763213   
## converged  
## trying - pitch   
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 448.953157  
## iter 20 value 229.383602  
## iter 30 value 217.811113  
## iter 40 value 214.815547  
## final value 214.744279   
## converged  
## Df AIC  
## - duration 10 447.0880  
## <none> 12 449.0205  
## - pitch 10 449.4886  
## - height 10 577.5264  
## - aircraft\_01 10 642.8404  
## - speed\_ground 10 1530.4118  
## # weights: 18 (10 variable)  
## initial value 915.144036   
## iter 10 value 392.072308  
## iter 20 value 227.605114  
## iter 30 value 218.318199  
## iter 40 value 213.555421  
## final value 213.544013   
## converged  
##   
## Step: AIC=447.09  
## y ~ aircraft\_01 + speed\_ground + height + pitch  
##   
## trying - aircraft\_01   
## # weights: 15 (8 variable)  
## initial value 915.144036   
## iter 10 value 376.398431  
## iter 20 value 313.337110  
## iter 30 value 312.671497  
## final value 312.576442   
## converged  
## trying - speed\_ground   
## # weights: 15 (8 variable)  
## initial value 915.144036   
## iter 10 value 758.193872  
## final value 757.747220   
## converged  
## trying - height   
## # weights: 15 (8 variable)  
## initial value 915.144036   
## iter 10 value 343.161878  
## iter 20 value 283.640109  
## iter 30 value 280.755767  
## final value 280.545060   
## converged  
## trying - pitch   
## # weights: 15 (8 variable)  
## initial value 915.144036   
## iter 10 value 410.465535  
## iter 20 value 230.928630  
## iter 30 value 219.703472  
## iter 40 value 215.915775  
## iter 50 value 215.915455  
## final value 215.915396   
## converged  
## Df AIC  
## <none> 10 447.0880  
## - pitch 8 447.8308  
## - height 8 577.0901  
## - aircraft\_01 8 641.1529  
## - speed\_ground 8 1531.4944

## Final Model

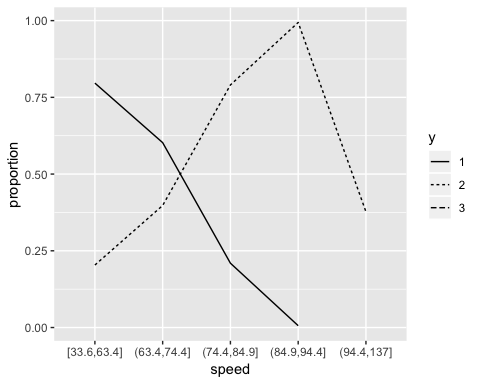
summary(step\_multi)

## Call:  
## multinom(formula = y ~ aircraft\_01 + speed\_ground + height +   
## pitch, data = faa\_multi)  
##   
## Coefficients:  
## (Intercept) aircraft\_011 speed\_ground height pitch  
## 2 -22.47102 4.110431 0.2485276 0.1485459 -0.2499589  
## 3 -142.27344 9.242677 1.2714499 0.4064769 1.2881498  
##   
## Std. Errors:  
## (Intercept) aircraft\_011 speed\_ground height pitch  
## 2 2.06743732 0.4228720 0.01999997 0.01733431 0.2638406  
## 3 0.03618892 0.8464824 0.02863608 0.03923758 0.7354845  
##   
## Residual Deviance: 427.088   
## AIC: 447.088

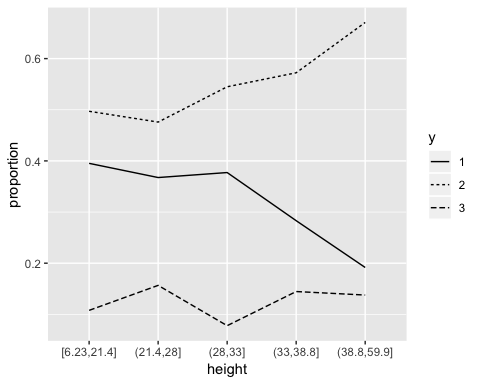
## 

## Plots

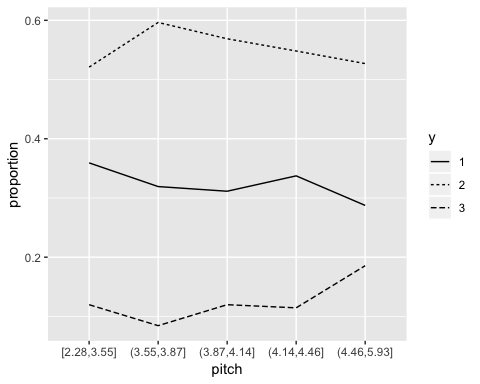
sg<-mutate(faa\_multi, speed=cut\_number(speed\_ground,5)) %>%  
group\_by(speed, y) %>% summarise(count=n()) %>%  
group\_by(speed) %>% mutate(etotal=sum(count),  
proportion=count/etotal)  
ggplot(sg, aes(x=speed, y=proportion, group=y,linetype=y)) + geom\_line()



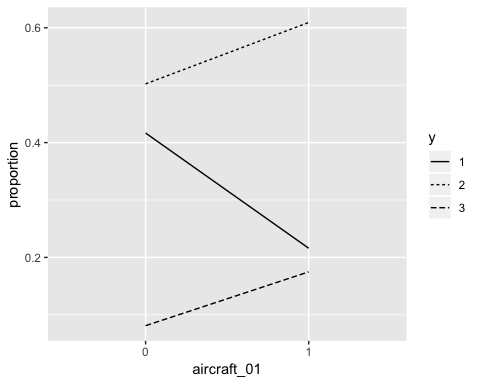
h<-mutate(faa\_multi, height=cut\_number(height,5)) %>%  
group\_by(height, y) %>% summarise(count=n()) %>%  
group\_by(height) %>% mutate(etotal=sum(count),  
proportion=count/etotal)  
ggplot(h, aes(x=height, y=proportion, group=y,linetype=y)) + geom\_line()



p<-mutate(faa\_multi, pitch=cut\_number(pitch,5)) %>%  
group\_by(pitch, y) %>% summarise(count=n()) %>%  
group\_by(pitch) %>% mutate(etotal=sum(count),  
proportion=count/etotal)  
ggplot(p, aes(x=pitch, y=proportion, group=y,linetype=y)) + geom\_line()



air<-group\_by(faa\_multi,aircraft\_01, y) %>%  
summarise(count=n()) %>% group\_by(aircraft\_01) %>%  
mutate(etotal=sum(count), proportion=count/etotal)  
ggplot(air, aes(x=aircraft\_01, y=proportion,  
group=y, linetype=y)) + geom\_line()



## Observations:

* Aircraft type,speed ground,height,pitch are the most influential on level of the landing distance
* From the above plots we can observe,the aircrafty typeBoeing has higher chances of either having very high landing distances or very low compared to Airbus.Also, the probabilty ratio of landing distance of level 3 and 1 will multiply by e^9 factor
* With pitch increasing the chances of risky landing increases and also obviously with speed ground

## Prediction

preds<-  
data.frame(faa\_multi$y,predict(step\_multi,data.frame(faa\_multi),type="probs"))  
xtabs(~predict(step\_multi)+faa\_multi$y)

## faa\_multi$y  
## predict(step\_multi) 1 2 3  
## 1 232 35 0  
## 2 37 420 6  
## 3 0 5 98

# 3. Predicting the number of passengers using Poisson regression

* Number of Passengers is a non-negative integer and is a count. Poisson Distribution might be good choice to explain Number of Passengers.
* For Poisson regression , original distance variable is used

poisson\_data<-FAA[,-c(1,6)]  
poisson\_data$duration<-ifelse(is.na(poisson\_data$duration),mean(poisson\_data$duration,na.rm=TRUE),poisson\_data$duration)  
  
pois\_reg<-glm(no\_pasg~.,family = "poisson",data=poisson\_data[,-6])  
  
step\_pois<-step(pois\_reg)

## Start: AIC=5730.25  
## no\_pasg ~ speed\_ground + height + pitch + duration + aircraft\_01  
##   
## Df Deviance AIC  
## - speed\_ground 1 780.13 5728.2  
## - pitch 1 780.27 5728.4  
## - aircraft\_01 1 780.33 5728.5  
## - duration 1 781.22 5729.3  
## - height 1 781.92 5730.0  
## <none> 780.13 5730.2  
##   
## Step: AIC=5728.25  
## no\_pasg ~ height + pitch + duration + aircraft\_01  
##   
## Df Deviance AIC  
## - pitch 1 780.27 5726.4  
## - aircraft\_01 1 780.33 5726.5  
## - duration 1 781.23 5727.4  
## - height 1 781.93 5728.0  
## <none> 780.13 5728.2  
##   
## Step: AIC=5726.39  
## no\_pasg ~ height + duration + aircraft\_01  
##   
## Df Deviance AIC  
## - aircraft\_01 1 780.66 5724.8  
## - duration 1 781.35 5725.5  
## - height 1 782.04 5726.2  
## <none> 780.27 5726.4  
##   
## Step: AIC=5724.78  
## no\_pasg ~ height + duration  
##   
## Df Deviance AIC  
## - duration 1 781.69 5723.8  
## - height 1 782.45 5724.6  
## <none> 780.66 5724.8  
##   
## Step: AIC=5723.81  
## no\_pasg ~ height  
##   
## Df Deviance AIC  
## - height 1 783.45 5723.6  
## <none> 781.69 5723.8  
##   
## Step: AIC=5723.57  
## no\_pasg ~ 1

summary(step\_pois)

##   
## Call:  
## glm(formula = no\_pasg ~ 1, family = "poisson", data = poisson\_data[,   
## -6])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.4606 -0.6629 -0.0082 0.6285 3.2550   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.095404 0.004471 916.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 783.45 on 832 degrees of freedom  
## Residual deviance: 783.45 on 832 degrees of freedom  
## AIC: 5723.6  
##   
## Number of Fisher Scoring iterations: 4

## Observations:

* Poisson regression is performed.
* From the step wise regression it can be observed that there is a only small variation in the AIC of different model.
* Height can be considered the most significant of all the variables to predict Number of Passengers